



# Toward Augmented Familiarity of the Audience with Digital Musical Instruments

Olivier Capra, Florent Berthaut, Laurent Grisoni

## ► To cite this version:

Olivier Capra, Florent Berthaut, Laurent Grisoni. Toward Augmented Familiarity of the Audience with Digital Musical Instruments. Music Technology with Swing, Springer, pp.558-573, 2018, 978-3-030-01691-3. hal-01954791

**HAL Id: hal-01954791**

**<https://hal.science/hal-01954791>**

Submitted on 3 Jun 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Toward Augmented Familiarity of the Audience with Digital Musical Instruments

Olivier Capra<sup>1</sup>, Florent Berthaut<sup>1</sup>, and Laurent Grisoni<sup>1</sup>

Univ. Lille, INRIA, CRIStAL, France

`olivier.capra@etu.univ-lille1.fr`

`florent.berthaut@univ-lille1.fr` `laurent.grisoni@univ-lille1.fr`

**Abstract.** The diversity and complexity of Digital Musical Instruments often lead to a reduced appreciation of live performances by the audience. This can be linked to the lack of familiarity they have with the instruments. We propose to increase this familiarity thanks to a trans-disciplinary approach in which signals from both the musician and the audience are extracted, familiarity analyzed, and augmentations dynamically added to the instruments. We introduce a new decomposition of familiarity and the concept of *correspondences* between musical gestures and results. This paper is both a review of research that paves the way for the realization of a pipeline for augmented familiarity, and a call for future research on the identified challenges that remain before it can be implemented.

**Keywords:** digital musical instrument, performance, familiarity, causality, augmentations, correspondences

## 1 Introduction

Whether as a musician or as a spectator, the experience of live music has very particular characteristics. It is often immersive, as intimate as collective and implies different modalities with both low and high level cognitive engagements. Moreover, a musical experience is often a corporeal, aesthetic as well as emotional commitment and is therefore difficult to define. Despite their great diversity and the parallel treatments they require, we assimilate all these aspects without any apparent effort. We naturally build this intense feeling that we all have experienced by attending a concert. Beyond the unfathomable subjective part of this musical experience (i.e. tastes and colors), we can identify objective characteristics that influence how we perceive live music, and among these, the way musicians interact with their instrument and how we perceive and integrate these interactions.

On the one hand, in the production of music with acoustic instruments, gestures and sounds are intrinsically linked by the laws of physics or at least by intuitive connections that do not need any prior explanation. Beside the music they produce, every interaction is visible. Thereby the expressive intentions, as well as the intensity of emotions, are particularly vivid and underline the role of

a multimodal integration [42]. Along this line, the degree of perceived control, influenced by the additional information usually extracted from the musician’s interactions, participates in the emergence of the liveness of the performance.

On the other hand, the experience of live music produced by digital musical instruments (DMIs) may suffer from a deterioration of this intuitive link between the behavior of the musician, the inner mechanisms of the instrument and the sounds actually produced. When what we hear is not directly linked to what we see or what we may infer from a given gestural behavior, when music is not the direct and causal consequence of the specific gestures of the musician, then we may lose the multimodal integration that contributes to the immersive experience of live music.

These considerations have led to an increasing amount of recent researches on measuring the audience experience [2, 26, 18, 19, 17, 12, 3, 10], sometimes used as a way to evaluate the instruments themselves [44]. The less familiar we are with an instrument, the less we are able to perceive the fine relationship musicians build with their instrument. This lack of coherence in multimodal information might downgrade the attributed agency [7], that is the perception of how much the musician is controlling their performance. Thus the knowledge we have about a DMI or its obvious behavior are crucial features to understand and integrate the interactions and thus fully experience live music. How do we ensure familiarity with DMIs ?

### 1.1 Improving familiarity with Digital Musical Instruments

Familiarity is the feeling of knowing about the behavior and the possibilities of an instrument. It can be associated with other concepts in the literature as *mental model* [18] or as an extension of *transparency* [16]. We believe that familiarity is a strong component of the live music experience as it provides the spectator the ability to detect the intentions and the virtuosity of the performance of the musicians as well as their errors. Several attempts have been made at improving the familiarity. In this section, we propose a quick review of these solutions and their limitations.

**Building a repertoire** Quite naturally, a first way to increase the familiarity of people with a particular instrument is to promote its use and dissemination by building a repertoire of compositions around it. Once a majority of spectators have seen the instrument played by many musicians in different contexts, or even practiced it themselves, they are aware of its potential for musical expression and of its behavior. Thereby, the familiarity issue does not exist anymore. This kind of “natural familiarity” is definitely effective but requires a large amount of time and energy to be achieved. Moreover, this method is not compatible with the very idea behind DMIs and the exponential creativity they embed. Whether a musician wants to evolve its instrument or let other musicians modify it, as soon as the instrument changes, the whole process of natural familiarity needs to start all over again.

**Explaining the instruments.** When the problem of familiarity lies partially in the understanding of the operations of an instrument, demonstrating the behavior of the instrument can be a valuable solution. Building the familiarity with a pedagogical method is a simple way to make an audience understand what is going on on stage. Before, after and even in breaks during the performance, the musicians can explain how their instrument works. Prior hands-on demos, where the audience can actually play the instrument, are also a good way to increase the familiarity. However, both these methods trigger some reservations as the technical understanding of an instrument is not necessarily linked to a better appreciation of the performance. For example, Bin et al. [11] demonstrate that explanations before the performance do improve the understanding of the instruments and its mechanisms but do not increase either the appreciation or interest. Besides these results, the audience may forget or unconsciously misreport important details from the performance [29] and the demos may not be possible when dealing with a large audience.

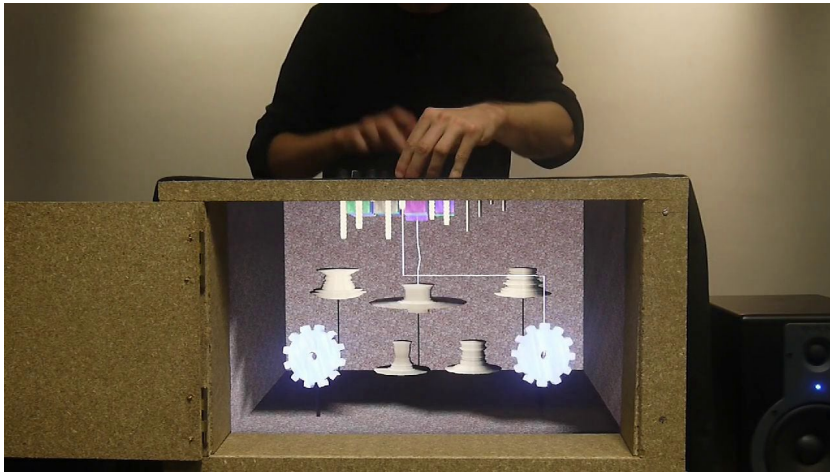
**Designing for transparency.** Another strategy to increase the experience of spectators watching a musician playing an instrument they have never seen before lies in the design of the instrument itself. The idea is to expose a clear link between the gestures and the sound modifications operated by these gestures, what is called the *mapping* of an instrument. The fluency of perception of these mappings is often called “transparency”, and is defined by Fels et al. [16] as follows :

For the lay audience, this understanding is derived from cultural knowledge, including percepts of physical causality relationships.

Aiming for transparency from the very first steps of the design of a musical device can lead to “easier to perceive” instruments [31]. Following that lead, an interesting way to increase the transparency is the use of metaphors [16]. In that case, the common background of inexperienced audience, their general knowledge, are used as a mold to grow new knowledge about an instrument mapping. For instance, the timbre of a sound can be modified as the shape of an associated graphical representation that is getting sculpted. However, the instruments designed according to these specifications tend to maintain the familiarity we have with acoustic instruments and physical laws. Thus, despite the gain in transparency, this method may narrow the design possibilities usually available for DMIs.

**Visually augmenting the instruments.** Berthaut et al. [7, 8] proposed to increase the familiarity by recreating a link of causality through visual augmentations of the instruments. Based on the attribution of causality by spectators (attributed agency) to a music performance, the visual augmentations provide insights on the musician’s interactions with its instrument. Animated 3D objects overlapping the device reveal the relationship between the gestures and the

musical result to the audience. These augmentations are based on Wegner’s criteria of apparent mental causation [43] (See 2.1). By exposing the details of the interactions, the setup contributes to re-link the gestures of the artist to their intentions and expressiveness. Besides, the augmentations enable the distinction between automated part and actual live music production. Such a discrimination between live and pre-recorded music appears to be crucial in the experience of live music precisely because it is supposed to be a live performance. However, these augmentations do not take into account the audience reaction or their expertise. Indeed, the familiarity with an instrument is a personal characteristic that cannot be generalized. A concert in front of specialists or naive people does not imply the same requirement of explanations or augmentations.



**Fig. 1.** Visual augmentations of DMIS for the increase of familiarity. The inner mechanisms of the instrument are dynamically revealed to the audience. (From Berthaut et al. 2013 [8])

**Toward a new method.** These different methods offer a rather effective way to increase the understanding of the behavior of DMIs. However they don’t fully cover a major aspect of the music experience which is the real time multimodal integration of visual and auditory cues. Furthermore, while the visual augmentations gave good results they do not take into account the audience expertise that may appeal for specific levels of details in the augmentations. Moreover, in some cases, non-visual augmentations may be more relevant. Thereby, we propose new insights to improve the familiarity in a more adaptable and reliable way.

## 1.2 Contribution

Our contribution is two-fold. First, we propose a conceptual pipeline to improve the familiarity of the audience with DMIs in real-time by analyzing spectators' reactions, musicians' actions, musical outputs and by augmenting the instrument. To do so, we introduce the concept of *correspondences*. Second, for each module of this pipeline, we briefly review relevant knowledge in computer and cognitive sciences. We also identify interdisciplinary challenges that need to be solved to reach a functional software implementation of the pipeline.

## 2 Toward augmented familiarity

In this section we propose a novel approach for improving familiarity, formalized by a software pipeline. While it remains conceptual, it was designed so that it could be implemented by addressing the challenges described in Section 2.2.

Here is a scenario that we envision with our pipeline :

*Patricia attends an electronic music concert. At the entrance, she is given a small device equipped with physiological sensors (a choice of either a bracelet or a special glass that she holds). During the concert, she has trouble understanding what is happening, in particular what the musician's action on the sound is. The device senses a change in a set of physiological signals, that corresponds to a loss in familiarity, and sends the data to a server. Patricia may also directly indicate her loss in familiarity with a graphical slider on an app on her smartphone. Simultaneously, this server has been analyzing the musician's gestures, the flow of data inside the instrument and the musical output. When it receives Patricia's familiarity signals, the server, with settings defined by the musician, selects the adequate augmentations to be displayed. They aim at compensating for the familiarity disruption caused by the musical interaction context. Consequently, visual augmentations are displayed around the musician either for Patricia alone when she watches the performance through her smartphone (e.g. with video augmented-reality), or for the group of people around her using a mixed-reality display. They provide information that improve her degree of familiarity, allowing her for example to perceive the link between the musician's gesture and the resulting sound, and to enjoy the performance to a larger extent.*

### 2.1 General approach

To improve the liveness of music performances with DMIs, we suggest increasing the familiarity thanks to a trans-disciplinary approach in which both human and technological signals are analyzed. To achieve this goal, we first clarify the notion of familiarity. We then propose the concept of *correspondences*, which describes musical interactions and the way they are perceived. Finally we present a potential pipeline that would extract and process signals from the musician and from the audience to compose augmentations.

**Familiarity dimensions.** As other high-level cognitive abilities, familiarity is not a unitary notion. It can be decomposed into components that differ by their nature, their inner rules as well as the type of information they convey. Besides, familiarity relates to expertise or attention and thus impacts on brain and motor system activities. We will describe in Section 2.2 how previous research shows that these activities can be detected in more detail. Familiarity may also relate to various parts of the musical interaction that is perceived. Spectators can be more or less familiar with : musical gestures, relations between gestures and sound, instruments capabilities or musical genres. All these factors contribute to the general familiarity one has with an instrument. Following research by Fyans et al. [18, 19] and Belloti et al. [5], Barbosa et al. [2] evaluate the audience experience through their comprehension of cause, effect, error, intention and mapping. However, measuring the familiarity might not be sufficient to inform the design of instruments or to augment performances if the origin of this familiarity is not understood correctly. For example, the perception of expertise, skill, error or intention of the performer, are consequences of the familiarity one has with a musical interaction rather than dimensions of the familiarity itself. For instance, the more familiar we are with an instrument, with the way the sound is produced and with the context of the performance, the better we can assess the quality of the performance and perceive the errors and the intention of the musician. Therefore while they can help measure the familiarity, to which they are correlated, these features do not provide insights on the reasons for this familiarity. One must therefore investigate the features which, when changed, have an impact on the overall familiarity, and in turn on the perception of expertise, skill, error or intention.

Taking this into account, our approach is to decompose familiarity into dimensions that we can both evaluate and improve. Based on results in HCI, NIME, and cognitive sciences that have explored various aspects of familiarity, we propose to decompose it into five components : *Causality*, *Instrumentality*, *Instrument expertise*, *Musical culture* and *Musical genre expertise*.

These in turn contain dimensions that can be evaluated independently :

*Causality* relates to the apparent mental causation, e.g. the judgment of causality of one's action, defined by Wegner et al [43], where each of the three following dimensions are required to establish a judgment of causality of one's action:

- *Priority* - The thought should precede the action at a proper interval
- *Consistency* - The thought should be compatible with the action
- *Exclusivity* - The thought should be the only apparent cause of action

The consistency aspect is also closely related to the notion of naive physics, that is a set of innate or common knowledge about the physical world like gravity or friction that can also be used in digital interfaces as physical metaphors [21].

As proposed in [7], we transpose this model from the perception of one's own intentions/actions to the perception of others' actions. By doing so, we extend the concept of agency to the rather recent concept of *attributed agency*.

*Instrumentality* relates to common knowledge that allows one to predict the range of sound possibilities of a musical instrument from its appearance. It is composed of two dimensions : the *composition*, i.e. shape and material, and *behavior*, i.e. mechanisms and degree of autonomy from the musician’s actions.

*Instrument expertise* relates to the exposure the spectator has had to the instrument, from a first-time observer to an expert player.

*Musical culture* corresponds to a basic knowledge of musical theory, that can for example be used to represent pitch as vertical position of a graphical element, i.e. mimicking a staff.

*Musical genre expertise* is composed of dimensions that correspond to the familiarity with the specifics of a genre, such as of *structure* and *constraints*. The effect of this category is obvious in the study by Bin et al. [11], where the same instrument played in two genres has a different impact on familiarity.

Notice that each component has a specific weight on the overall process. Furthermore, this model is consistent with the partial results that previous strategies have achieved. *Culture and expertise* for the “building a repertoire” and the “prior explanations” strategies, *Instrumentality* and *Culture* for “the design for transparency” and *Causality* for the augmentation strategy. This decomposition might evolve according to findings from the implementation phase. Some dimensions could prove harder to evaluate or others may emerge.

Finally, the complete decomposition of familiarity that we propose includes the following dimensions :

$$\begin{aligned}
 \text{Familiarity} = & C_{\text{exclusivity}} + C_{\text{priority}} + C_{\text{consistency}} \\
 & + I_{\text{composition}} + I_{\text{behavior}} \\
 & + E_{\text{expertise}} \\
 & + M_{\text{culture}} \\
 & + G_{\text{structure}} + G_{\text{rules}}
 \end{aligned} \tag{1}$$

**Correspondences.** In order to handle the heterogeneous data (physiological, behavioral, musical, visual, mechanical) associated with the dimensions of familiarity, we introduce the notion of ***correspondence***: a conceptual object that stands as a digital multidimensional representation of a musical interaction. This object associates several *properties* : musical output (audio), gesture data (time series of 3D coordinates), visual data (video of the movement), control data (time series of sensor values) and the *source* of the interaction. This last property specifies the author of the interaction. Indeed, interactions are not only produced by one musician but can be triggered and mastered by another one in a collaborative performance or, more usually, by autonomous prepared processes, e.g. automations or playlists.

Each property is subdivided in three elements : a *raw format element*, a *semantic element*, and a *classification element*.



The *raw format element* is a pointer to a collection of related files. For instance, the raw element of the visual property of a clap correspondence is a pointer to a collection of short videos showing a clap from different angles and velocities and the raw element of the audio property a pointer to a collection of short audio recordings of a clap.

The *semantic element* is composed of annotated descriptors of the property. For our clap example, the semantic element of the visual property could list obvious descriptors as "hands", "clap", "applause", "brief" but also more precise descriptors relative to specific taxonomies developed in the different analysis of gesture and sound.

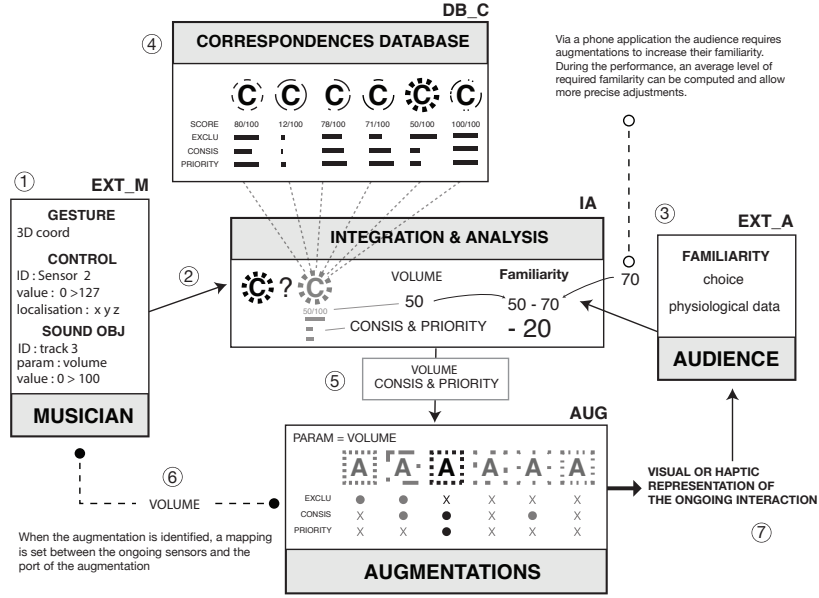
The descriptors can be freely provided by the audience or taken from refined existing models such as [13] or [22]. For instance, "excitation gesture" or "effective gesture" could be used to populate the semantic element of the gesture property of a correspondence. Interestingly, by holding descriptors for each property of a unique interaction (aka correspondence), the semantic element offers the possibility to analyze the links between descriptors from different properties by the analysis of their co-occurrence. Finally, looking forward to the implementation, the processing of meaningful descriptors can rely on the strong foundation of web semantics [6].

The *classification element* is dedicated to the classification of the raw elements through machine learning methods and thus allows for the pairing of close correspondences with respect to their different properties. To compute such a distance between correspondences, multidimensional vectors can be used alongside more specific techniques as neural networks, informed by an efficient semantic analysis thanks to the semantic element of a given property.

In addition, correspondences hold a *score* for each of the familiarity dimensions described earlier (causality, instrumentality, ...). For example, two correspondences with the same musical result, e.g. the fade-in of an audio loop, can have different gesture and control properties depending on the mapping chosen for the DMI. While a continuous gesture on a fader would have a high score (meaning, the most natural way to fade) for the *consistency* dimension, a discrete tap on a pad would have a low one, since the effect would no be consistent with the cause for a spectator, e.g. discrete input and continuous output. This score of familiarity held by each correspondence plays a central role in our approach for an increase of familiarity as it allows for sorting correspondences from obvious to abstruse for a given audience.

Finally, as an interaction may be more complex and decomposed in a sequence of more basic interactions, a correspondence can also be linked to other, simpler correspondences.

**Pipeline.** Our envisioned pipeline (Fig 2) is composed of five modules that handle the extraction of the data from the musician and their instrument (*EXT\_M*), the extraction of physiological and subjective data from the audience (*EXT\_A*), the processing of the data (*IA*) and the selection of fitting augmentations (*AUG*)



**Fig. 2.** The proposed pipeline with its main modules and data flows

based on a database of correspondences (*DB\_C*). The pipeline is used at three different moments : before, during and after a performance.

Before the performance, correspondences, coming from a shared online database, or recorded specifically for the instrument, are saved in the *DB\_C* module.

① During the performance, *IA* receives musical interaction data extracted by *EXT\_M*. This data consists of both dynamic values such as gestural parameters, audio features and control values which will be used directly by *IA*, and of physical (position of gestures and sensors), logical (tracks, effects, synthesizers, ...) and structural information about the musical interaction, which will be used by *AUG*. ② *IA* builds and manages a set of ongoing correspondences from these signals. ③ Simultaneously, *IA* receives the familiarity evaluation (with the identification of the sensed individual or group) from *EXT\_A*. If the familiarity evaluation is low, ④ *IA* finds in *DB\_C* correspondences similar to the ongoing one, and select the familiarity dimensions that need to be compensated depending on their scores in these. *IA* then sends *AUG* the data required for the augmentation : live correspondences, associated signals from *EXT\_M* with dynamic and structural data and identification of the source of the familiarity evaluation (in order to display the augmentations only to the correct person or group). ⑤ *AUG* creates (or selects if it already exists) the augmentation that matches the received structural data, for example a visual augmented-reality link between the physical position of a sensor and a virtual representation of an audio track. ⑥ Augmentations are then connected to *EXT\_M* and listen to the

signals required to update the augmentation, e.g. control values for the sensor, loudness of the track. After the performance, the familiarity extracted from the audience can be reused to refine the scores for the familiarity dimensions in each detected correspondence of *DB\_C*.

## 2.2 Modules

**Database of correspondences (*DB\_C*).** *DB\_C* manages all pre-existing correspondences. These can be generic, or specific to an instrument or performance. It receives queries from *IA* to select correspondences matching the ones detected during the performance. A very promising approach in dealing with heterogeneous data is the use of databases. Even if the efficiency of machine learning and analysis tools is still evolving, a numerous amount of initiatives, especially in analysis of emotion (DEAP[24], RECOLA[35], EATMINT [15]) but also in music-related actions [20], contribute to shared databases that compile multi-modal and synchronized experimental data. The main goal of these databases is to predict complex and abstract states, e.g. the emotional state of an individual, thanks to the analysis of their physiological and behavioral signals such as face expression, electrodermal activity or heart rate variation. Most of the existing databases are composed of 15 to 30 entries referencing data of diversified nature. To allow the gathering of a more relevant amount of cross data, we first need to facilitate the indexing thanks to the more and more intuitive and effective front-end technologies of the web. We propose to develop a web interface that could provide the specific tools to aggregate the data required to constitute the properties of a correspondence. A typical correspondence would require a short video footage of a gesture, the motion capture of this gesture, sensor value and audio output of the instrument. Tags could also be manually added for each of these properties. The interface could be accessible from an open web platform where artists as well as researchers could populate the database, to constitute their own correspondences and therefore optimize their pipeline with more personal choices of gestures, mappings or sound processes. The same online platform can then be used for crowdsourced online evaluation sessions in which correspondences are exposed to participants with different levels of expertise. Their task is to indicate their understanding, using a questionnaire along the familiarity dimensions, and tag the correspondences.

**Extraction of the musician’s interactions (*EXT\_M*).** As explained in Section 2.1, *EXT\_M* extracts data from the instrument and musician’s gesture which is then sent to *IA* in order to be aggregated into *live correspondences*. At the instrument level, *EXT\_M* extracts sensors states, mapping values as well as musical result of the musician interactions. At the musician level, *EXT\_M* extracts the control gestures, the body movements and physiological signals. In addition to these signals, structural information needs to be provided for further use in the augmentations, such as the position of the physical sensors of the DMI and the position of the musician’s hands, the list of tracks, effects or other

sound processes with their names, or the mappings between sensors and sound parameters. While some of these obviously need to be defined by the musician manually, or sensed by devices external to the instrument, others can be extracted through a trans-disciplinary approach. Regarding the instrument input, research on gesture recognition, especially concerning hand gestures as demonstrated by Rautaray & Agrawal [34], can be used to identify the performed musical gestures. Regarding the instrument output, research in music information retrieval (MIR) provides tools for segmenting music from the audio signal only [32] using spectral, tonal, rhythmical descriptors and methods. Finally, the extraction can be facilitated by a multimodal approach, similarly to techniques developed for video indexing [40], as some events might appear clearer on one channel than on others. For example, control data might inform on the temporal boundaries of a change in the sound that can then be analysed

We identify two main challenges for the implementation of this module. The first is the access to the data from the DMIs. In fact, while extraction from the audio signal provides many features that can be used to detect correspondences, it might not be enough for precise analysis. In order to access pre-mappings and post-mappings data, to differentiate between manual and automated changes and to analyze the output of individual tracks or other sound processes, one can not rely solely on the DMIs inputs and outputs, i.e. additional software components will need to be integrated. In most DMIs, plug-ins can be added at various stages of the instrument. However, the API might not provide enough information on the instrument to a single plugin. For example, one plugin per track might be needed to access and send the audio output features for each separately. The integration of *EXT\_M* will be simpler if DMIs are built using patching environments such as PureData or Max/MSP, where the musician has more control over the architecture of the instruments. The second challenge is to combine detailed but costly and slow analysis of features for the detection of correspondences with maximum accuracy, and fast but less accurate analysis of features for the update of augmentations in *AUG*.

**Extraction of the audience familiarity (*EXT\_A*).** The role of this module is to extract the audience subjective and objective information in order to inform *IA*.

It is now commonly accepted that the live music experience engages multiple complementary processes of low level perception, en-action and embodiment ([17], see Leman & Maes for a review [28]), processing of hierarchical and sequential information [30, 25], as well as strong affective and social aspects. This perspective of complementarity is also included in recent studies relative to music produced with DMIs [37]. The evaluation of such a rich experience triggers multiple methodological difficulties. Subjective assessment is therefore a common method that can be conducted through post-performance questionnaires [27, 44, 12], emotion rating during the concerts [38, 12] or when viewing an audio-video recording of a performance [4].

Post-performance questionnaires are a good source of information but may be less precise than live subjective reports. In addition, there are initiatives that do not require the participants to directly communicate their introspective evaluation. These more objective measurements require very specific equipment as eye trackers [4] and, to date, suffer from a lack of out-of-the-lab physiological measurements. Familiarity is one of the key aspects of the experience of live music and, as its other dimensions, relies on multiple underlying mechanisms. Therefore, as the mentioned studies above, and to initiate ecological (i.e. out of the lab, in “real life”) measurements of physiological signals, we propose a dual methodology to extract the familiarity of the audience. First, we base the subjective assessment on a familiarity application for mobile phone. Its main purpose is to supply *IA* with data about the ongoing familiarity of the audience, from a graphical familiarity slider that spectators activate.

As other authors [3][12], we believe this continuous survey could be a good methodological answer to the reservations we exposed about questionnaires after the performance. The second part of the extraction relies on physiological signals. Neuroscience studies show that the expertise, a key component of familiarity, influences the perception of action [14]. Those findings, applied to music expertise, may lead to a better understanding of its role in the live music experience. By measuring the peripheral signals, we aim at discovering potential patterns that could correlate with the subjective data we extract. Widely used in the emotion studies, and rather suitable to extract in natural condition, four signals are particularly interesting in our musical context : Heart rate variability [36], electrodermal activity [39] and oculometry (eye tracking + pupillometry). The complementary analysis of extracted features of these signals already gave interesting result in the emotion classification by machine learning algorithms [23] and need to be further extended in music experience studies. Moreover, these signals can already be acquired by wearable devices and the quantified self movement [41] will surely provide more accurate and affordable devices in the near future.

**Integration and Analysis (*IA*).** *IA* is the central hub that connects to all the other modules. Its role is to : 1) compute the live correspondences with the data extracted from the musician and their instrument; 2) match the computed live correspondence with a correspondence from the database to calculate the familiarity dimensions that need to be augmented; 3) supply the *AUG* module with the information needed for the relevant augmentation selection. To fulfill these tasks, the module can rely on the classification element of each property of correspondences. Alongside the raw and the semantic parts, the classification element is a machine learning model dedicated to the classification of raw data. This model is pre-computed with the raw files registered in the correspondence. Its goal is to discriminate new stimuli and detect those who match with the recorded one. Considering the heterogeneity of the modalities, each property might require a specific machine learning model and specific extracted features. For example, MIR descriptors for a sound element and a deep convolutional

network for picture classification. The main idea is to use the set of models as a global digital representation of the correspondence that can either be projected, depending of the context, on a single and more easy to handle property or be represented as a multidimensional vector that allows similarity comparison of whole correspondences (the matching process). Considering the variety of data and processes it has to handle, this module needs to be regularly updated with recent findings in signal processing, machine learning or movement models. Without a strongly modular structure of the available tools that the pipeline has at its disposal, the framework may not be able to evolve and thereby join a long list of deprecated initiatives.

**Augmentation (*AUG*).** This module manages both a database of available augmentations and a set of active ones. When a correspondence needs to be augmented, *AUG* receives the data required to create a new or select an existing augmentation, such as the physical position of the gesture and sensors and logical components of the instruments, the familiarity dimensions that need to be compensated, the destination of the augmentation (individual or group) and the data from *EXT\_M* that the augmentation should listen to. The augmentations are selected from a database of augmentations designed to compensate the various dimensions of familiarity.

A number of research have shown the opportunities opened by augmentations to provide information on DMIs to the audience. For example, Perrotin et al. [33] used visualizations, projected on a screen behind an ensemble of DMIs, to help the audience understand better the contribution of each musician. In previous work, we proposed an augmented-reality (AR) approach where the visualizations are perceptually consistent with the physical instrument, e.g. visual links attached to the physical sensors. We designed a display where multiple spectators can reveal the augmentations [9] and all perceive them consistently. Finally, we proved the effect of 3D visual augmentations on the causality aspect of familiarity [7], i.e. the degree of control perceived by the audience. On the haptic side Armitage [1] has experimented with using haptic feedback to provide information on musician’s interaction during a live-coding performance. We believe that the first challenge is the creation of a framework that allows one to design augmentations according to the specific dimensions of familiarity that they compensate. Rules will need to be defined so that one can adapt an augmentation to the artistic specificity of a particular performance. A second challenge is the design of augmentations that provide just enough content to fill the multimodal gap without distracting the audience from the musical performance because of a too heavy cognitive load.

### 3 Conclusion

In this paper, we presented a novel approach for augmenting familiarity of the audience with Digital Musical Instruments and reviewed associated research results and challenges.

Among the perspectives, our first future work will be the construction of the initial database of correspondences that will help refining our familiarity decomposition.

The first ones will therefore be generated through a systematical process, i.e. by recording unitary interactions (e.g. one gesture one change in the sound) with variations on each dimension perceived by the audience, e.g. gesture type and amplitude, mapping, sound parameter, point of view, and so on.

In addition, we plan on investigating the impact of familiarity variations on electrophysiological signals from the audience, with the aim of finding realtime measurements that could be used to assess familiarity during performances and guide the choice of augmentations. This transdisciplinary approach could reveal significant research avenues especially with the democratization of wearable devices. We hope that this paper will trigger exciting new research in NIMes, and both cognitive and computer sciences.

## References

1. J. Armitage. Revealing timelines: Live coding and its gestures. In *Proceedings of ICLC*, 2016.
2. J. Barbosa, F. Calegario, V. Teichrieb, G. Ramalho, and P. McGlynn. Considering audience’s view towards an evaluation methodology for digital musical instruments. In *Proceedings of NIME*, 2012.
3. M. A. Baytas, T. Göksun, and O. Özcan. The perception of live-sequenced electronic music via hearing and sight. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, volume 16 of 2220-4806, pages 194–199, Brisbane, Australia, 2016. Queensland Conservatorium Griffith University.
4. M. A. Baytas, T. Göksun, and O. Özcan. The perception of live-sequenced electronic music via hearing and sight. In *Proceedings of NIME*, 2016.
5. V. Bellotti, M. Back, W. K. Edwards, R. E. Grinter, A. Henderson, and C. Lopes. Making sense of sensing systems: five questions for designers and researchers. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 415–422. ACM, 2002.
6. T. Berners-Lee, J. Hendler, O. Lassila, et al. The semantic web. *Scientific american*, 284(5):28–37, 2001.
7. F. Berthaut, D. Coyle, J. Moore, and H. Limerick. Liveness Through the Lens of Agency and Causality. In *Proceedings of NIME*, 2015.
8. F. Berthaut, M. T. Marshall, S. Subramanian, and M. Hachet. Rouages: Revealing the Mechanisms of Digital Musical Instruments to the Audience. In *Proceedings of NIME*, 2013.
9. F. Berthaut, D. Martinez Plasencia, M. Hachet, and S. Subramanian. Reflets: Combining and Revealing Spaces for Musical Performances. In *Proceedings of NIME*, 2015.
10. S. A. Bin, N. Bryan-Kinns, A. McPherson, et al. Hands where we can see them! investigating the impact of gesture size on audience perception. International Computer Music Conference, 2017.
11. S. A. Bin, N. Bryan-Kinns, and A. P. McPherson. Skip the pre-concert demo: How technical familiarity and musical style affect audience response. In *Proceedings of NIME*, 2016.

12. S. A. Bin, F. Morreale, N. Bryan-Kinns, and A. P. McPherson. In-the-moment and beyond: Combining post-hoc and real-time data for the study of audience perception of electronic music performance. In *IFIP Conference on Human-Computer Interaction*, pages 263–281. Springer, 2017.
13. C. Cadoz and M. M. Wanderley. Gesture-music, 2000.
14. B. Calvo-Merino, D. E. Glaser, J. Grèzes, R. E. Passingham, and P. Haggard. Action observation and acquired motor skills: an fmri study with expert dancers. *Cerebral cortex*, 15(8):1243–1249, 2005.
15. G. Chanel, M. Bétrancourt, T. Pun, D. Cereghetti, and G. Molinari. Assessment of computer-supported collaborative processes using interpersonal physiological and eye-movement coupling. In *Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on*, pages 116–122. IEEE, 2013.
16. S. Fels, A. Gadd, and A. Mulder. Mapping transparency through metaphor: towards more expressive musical instruments. *Organised Sound*, 7(2):109–126, 2002.
17. A. C. Fyans and M. Gurevich. Perceptions of skill in performances with acoustic and electronic instruments. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 495–498, Oslo, Norway, 2011.
18. A. C. Fyans, M. Gurevich, and P. Stapleton. Where did it all go wrong ? a model of error from the spectator’s perspective. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 171–172, Pittsburgh, PA, United States, 2009.
19. A. C. Fyans, M. Gurevich, and P. Stapleton. Examining the spectator experience. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 451–454, Sydney, Australia, 2010.
20. R. I. Godøy, A. R. Jensenius, A. Voldsund, K. H. Glette, M. E. Høvin, K. Nymoen, S. A. v. D. Skogstad, and J. Tørresen. Classifying music-related actions. 2012.
21. R. J. Jacob, A. Girouard, L. M. Hirshfield, M. S. Horn, O. Shaer, E. T. Solovey, and J. Zigelbaum. Reality-based interaction: A framework for post-wimp interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI ’08*, pages 201–210, New York, NY, USA, 2008. ACM.
22. A. R. Jensenius, M. M. Wanderley, R. I. Godøy, and M. Leman. Musical gestures. *Musical gestures: Sound, movement, and meaning*, 12, 2009.
23. J. Kim and E. André. Emotion recognition based on physiological changes in music listening. *IEEE transactions on pattern analysis and machine intelligence*, 30(12):2067–2083, 2008.
24. S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras. Deap: A database for emotion analysis; using physiological signals. *IEEE Transactions on Affective Computing*, 3(1):18–31, 2012.
25. E. Kohler, C. Keysers, M. A. Umiltà, L. Fogassi, V. Gallese, and G. Rizzolatti. Hearing sounds, understanding actions: action representation in mirror neurons. *Science (New York, N.Y.)*, 297(5582):846–8, 2002.
26. C.-H. Lai and T. Bovermann. Audience experience in sound performance. In *Proceedings of NIME*, 2013.
27. C.-H. Lai and T. Bovermann. Audience experience in sound performance. In *NIME*, page 170–173, 2013.
28. M. Leman and P.-J. Maes. The role of embodiment in the perception of music. *Empirical Musicology Review*, 9(3-4):236–246, 2014.
29. E. F. Loftus and J. C. Palmer. Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of verbal learning and verbal behavior*, 13(5):585–589, 1974.



30. I. Molnar-Szakacs and K. Overy. Music and mirror neurons: from motion to 'e'motion. *Social cognitive and affective neuroscience*, 1(3):235–241, 2006.
31. T. Murray-Browne, D. Mainstone, N. Bryan-Kinns, and M. D. Plumbley. The medium is the message: Composing instruments and performing mappings. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 56–59, 2011.
32. J. Paulus, M. Müller, and A. Klapuri. State of the art report: Audio-based music structure analysis. In *Proceedings of ISMIR*, 2010.
33. O. Perrotin and C. d'Alessandro. Visualizing gestures in the control of a digital musical instrument. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 605–608, London, United Kingdom, 2014. Goldsmiths, University of London.
34. S. S. Rautaray and A. Agrawal. Vision based hand gesture recognition for human computer interaction: a survey. *Artificial Intelligence Review*, 43(1):1–54, 2015.
35. F. Ringeval, A. Sonderegger, J. Sauer, and D. Lalanne. Introducing the recola multimodal corpus of remote collaborative and affective interactions. In *Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on*, pages 1–8. IEEE, 2013.
36. D. Sammler, M. Grigutsch, T. Fritz, and S. Koelsch. Music and emotion: electrophysiological correlates of the processing of pleasant and unpleasant music. *Psychophysiology*, 44(2):293–304, 2007.
37. J. C. Schacher and P. Neff. Skill development and stabilisation of expertise for electronic music performance. In *Proceedings of CMMR*. Springer, 2015.
38. E. Schubert, S. Ferguson, N. Farrar, D. Taylor, and G. E. McPherson. The six emotion-face clock as a tool for continuously rating discrete emotional responses to music. In *Proceedings of CMMR*, 2012.
39. H. Sequeira, P. Hot, L. Silvert, and S. Delplanque. Electrical autonomic correlates of emotion. *International journal of psychophysiology*, 71(1):50–56, 2009.
40. C. G. Snoek and M. Worring. Multimodal video indexing: A review of the state-of-the-art. *Multimedia tools and applications*, 25(1):5–35, 2005.
41. M. Swan. The quantified self: Fundamental disruption in big data science and biological discovery. *Big Data*, 1(2):85–99, 2013.
42. B. W. Vines, C. L. Krumhansl, M. M. Wanderley, I. M. Dalca, and D. J. Levitin. Music to my eyes: Cross-modal interactions in the perception of emotions in musical performance. *Cognition*, 118(2):157–170, 2011.
43. D. M. Wegner and T. Wheatley. Apparent mental causation: Sources of the experience of will. *American psychologist*, 54(7):480, 1999.
44. J. C. Wu, M. Huberth, Y. H. Yeh, and M. Wright. Evaluating the audience's perception of real-time gestural control and mapping mechanisms in electroacoustic vocal performance. In *Proceedings NIME*, 2016.